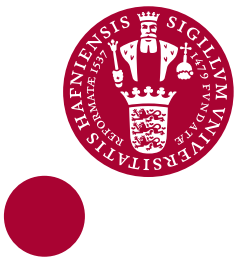


Discovery of -L-arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42

Viborg, Alexander Holm; Katayama, Takane; Arakawa, Takatoshi; Abou Hachem, Maher; Lo Leggio, Leila; Kitaoka, Motomitsu; Svensson, Birte; Fushinobu, Shinya

Published in:
Journal of Biological Chemistry

DOI:
10.1074/jbc.M117.792598

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

JBC Papers in Press. Published on October 23, 2017 as Manuscript M117.792598
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M117.792598

Discovery of α-L-Arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42

Alexander Holm Viborg[1,4,*], Takane Katayama[2,3], Takatoshi Arakawa[1], Maher Abou Hachem[4], Leila Lo Leggio[5], Motomitsu Kitaoka[6], Birte Svensson[4], and Shinya Fushinobu[1]**

From the [1]Department of Biotechnology, The University of Tokyo, Japan, [2]Graduate School of Biostudies, Kyoto University, Japan, [3]Faculty of Bioresources and Environmental Sciences, Ishikawa Prefectural University, Japan, [4]Department of Biotechnology and Biomedicine, Technical University of Denmark, Denmark, [5]Department of Chemistry, University of Copenhagen, Denmark, and [6]Food Research Institute, National Agriculture and Food Research Organization, Japan.

Running title: GH42 α-L-arabinopyranosidase subfamily identification

*Present address: Michael Smith Laboratories, The University of British Columbia, Vancouver, BC Canada V6T 124, Canada.
**To whom correspondence should be addressed: Prof. Shinya Fushinobu, Laboratory of Enzymology, Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan; Telephone/FAX: +81-3-5841-5151; E-mail: asfushi@mail.ecc.u-tokyo.ac.jp.

Keywords: bifidobacterium, carbohydrate metabolism, CAZyme, crystallography, β-galactosidase, glycobiology, microbiota, paeonolide, phylogenetics, probiotic

ABSTRACT

Enzymes of the glycoside hydrolase family 42 (GH42) are widespread in bacteria of the human gut microbiome and play fundamental roles in the decomposition of both milk and plant oligosaccharides. All GH42 enzymes characterized so far have β-galactosidase activity. Here, we report the existence of a GH42 subfamily that is exclusively specific for α-L-arabinopyranoside and describe the first representative of this subfamily. We found that this enzyme (*Bl*Arap42B) from a probiotic *Bifidobacterium* species cannot hydrolyze β-galactosides. However, *Bl*Arap42B effectively hydrolyzed paeonolide and ginsenoside Rb2, plant glycosides containing an aromatic aglycone conjugated to α-L-arabinopyranosyl-(1,6)-β-D-glucopyranoside. Paeonolide, a natural glycoside from the roots of the plant genus *Paeonia,* is not hydrolyzed by classical GH42 β-galactosidases. X-ray crystallography revealed a unique Trp345-X_{12}-Trp358 sequence motif at the *Bl*Arap42B active site, as compared to a Phe-X_{12}-His motif in classical GH42 β-galactosidases. This analysis also indicated that the C6 position of galactose is blocked by the aromatic side chains, hence allowing accommodation only of Ara*p* lacking this carbon. Automated docking of paeonolide revealed that it can fit into the *Bl*Arap42B active site. The Glc*p* moiety of paeonolide stacks onto the aromatic ring of the Trp252 at subsite +1 and C4-OH is hydrogen bonded with Asp249. Moreover, the aglycone stacks against Phe421 from the neighboring monomer in the *Bl*Arap42B trimer, forming a proposed subsite +2. These results further support the notion that evolution of metabolic specialization can be tracked at the structural level in key enzymes facilitating degradation of specific glycans in an ecological niche.

The complex microbial consortium in the human gastrointestinal tract, referred to as the gut microbiota, has an increasingly recognized vital role in human health (1). Human diet is rich in a wide variety of non-digestible saccharides from plants and animals, and glycan metabolism is a pivotal factor shaping the dynamics and the evolution of gut microbiota (2). Specific taxa gain a competitive edge in fitness in this highly competitive niche through metabolic

specialization, which is reflected by their enzymatic machinery being fine-tuned for specific glycans (3). Glycans containing β-galactoside are abundant in human infant and adult nutrition (4–6), and hydrolysis of the β-galactosidic bond by β-galactosidases is a prerequisite for their utilization.

Key enzymes in the hydrolysis and utilization of β-galactosides in gut microbes assign into glycoside hydrolase (GH) family 42 in the sequence-based classification system of the Carbohydrate Active enZymes (CAZy) database (7). Genomes of probiotic strains from the *Bifidobacterium* genus often encode several GH42 enzymes (8) with distinct subspecificities matching diversity and abundance of β-galactosides. For example, *B. longum* subsp. *infantis* encodes three GH42 β-galactosidases related to the utilization of the human milk oligosaccharide lacto-*N*-tetraose (LNT), β-galactosides from bovine milk and arabinogalactan, respectively (9). A similar specialization exists in *Bifidobacterium breve* that encodes two GH42 β-galactosidases, one targeting LNT (10) and the other plant galactan (11). Notably, *B. bifidum* utilizes mucin-derived β-galactosides from the glycoprotein layer coating the human host in epithelial colonocytes (12), which is proposed to be possible through a key GH42 β-galactosidase (13).

The genome of *B. animalis* subsp. *lactis* Bl-04 encodes two GH42 enzymes (14). Galacto-oligosaccharides (GOS) were shown to up-regulate gene locus balac_0848 (15), experimentally verified to be a β(1,6)/β(1,3) galactosidase (*Bl*Gal42A) (13). The second GH42 (*Bl*Arap42B) gene with locus tag balac_0053 was not differentially upregulated by GOS (15), and belongs to a distinct, not previously characterized clade of bifidobacterial GH42 enzymes (16).

This distinct clade is here shown to form a novel GH42 subfamily present throughout the bacterial kingdom, which seems to have exclusive α-L-arabinopyranoside specificity as evaluated for the soluble intracellular *Bl*Arap42B using a panel of *p*NP-substrates and oligosaccharides. *Bl*Arap42B effectively hydrolyzed paeonolide (Figure 1*A*), a plant glycoside which contains a non-reducing end α-L-arabinopyranoside and is found in the roots of the widespread plant genus *Paeonia* (17). This apparent metabolic specialization could be tracked at the structural level, as the crystal structure of *Bl*Arap42B suggested GH42 members with α-L-arabinopyranosidase activity recognize their substrates through an active site Trp345-Trp358 motif as compared to a Phe-His motif in classical GH42 β-galactosidases.

RESULTS

Biochemical properties–The recombinant *Bl*Arap42B was produced and purified to electrophoretic homogeneity. *Bl*Arap42B migrated as a single band in SDS-PAGE with the expected theoretically calculated molecular mass of 78.3 kDa. Analytical gel filtration data showed that the protein eluted as a single peak corresponding to 201 kDa. Although this is smaller than the theoretically calculated size for the trimer (235 kDa), GH42 enzymes, which are generally trimeric, usually exhibit similar discrepancy between their size estimations using gel filtration chromatography and crystallography (18). *Bl*Arap42B showed significant activity towards *p*-nitrophenyl (*p*NP) α-L-arabinopyranoside (-α-L-Ara*p*), very low activity towards *p*NP-β-D-galactopyranoside (*p*NP-β-D-Gal*p*) and *p*NP-β-D-fucopyranoside (*p*NP-β-D-Fuc*p*), and no detected activity on *p*NP-α-L-arabinofuranoside (*p*NP-α-L-Ara*f*), *p*NP-β-D-xylopyranoside (*p*NP-β-D-Xyl*p*), and *p*NP-β-D-glucopyranoside (*p*NP-β-D-Glc*p*) (Table 1). The kinetics of hydrolysis of *p*NP-α-L-Ara*p* were best described by Michaelis-Menten kinetics with k_{cat} = 160 ± 10 s^{-1} and K_m = 3.4 ± 0.4 mM. *Ri*Arap42B, a homologue from *Roseburia intestinalis* M50/1 (Firmicutes isolated from human gut), sharing 42% sequence identity with *Bl*Arap42B (Actinobacteria), has a similar activity profile on aryl substrates (Table 1). *Bl*Arap42B displays significant activity towards 2-acetyl-5-methoxyphenyl-α-L-Ara*p*-(1,6)-β-D-Glc*p* (paeonolide, Figure 1*A*) in a 30 min assay, with complete hydrolysis by overnight incubation as verified by TLC (Figure 1*B*). We also tested the activity of *Bl*Arap42B on plant glycosides containing a terminal α-L-arabinopyranoside group, paeonolide, ginsenoside Rb2, and quercetin 3-*O*-α-L-Ara*p* (Figure 1*A*), by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Figure 1*C*). Ginsenoside Rb2 is a triterpene saponin isolated from the root of *Panax ginseng* (19) while quercetin 3-*O*-α-L-Ara*p* is a flavonoid isolated

from the aerial parts (leaves) of *Alchemilla xanthochlora* (Lady's Mantle) (20) and *Psidium guajava* L. (guava) (21). *Bl*Arap42B hydrolyzed paeonolide and ginsenoside Rb2 (Figure 1*C*) but did not quercetin 3-*O*-α-L-Ara*p* (data not shown). Kinetic parameters toward the plant glycosides indicate that these substrates exhibit significantly lower K_m (<0.1 mM) and higher catalytic efficiencies (>1500 $mM^{-1}s^{-1}$) compared with those of *p*NP-α-L-Ara*p* (Table 2). By contrast, no activity was observed on an extensive series of β-galactosides from plants and mammals, including human milk (list of substrates in the experimental procedures). Bga42A (Blon_2016), Bga42B (Blon_2123), and Bga42C (Blon_2416) from *B. longum* subsp. *infantis* ATCC 15697, and *Bl*Gal42A from *B. animalis* subsp. *lactis* Bl-04, which are distantly related to *Bl*Arap42B (see Figure 4), did not cleave paeonolide, although they were able to slowly hydrolyze *p*NP-α-L-Ara*p* (Figure 1*B*).

Three-dimensional structure of Bl*Arap42B*–The three-dimensional structure of *Bl*Arap42B in its ligand-free form was determined by X-ray crystallography at 2.0 Å resolution with six chains (two trimers) in the asymmetric unit. Data collection, refinement and stereochemical statistics are summarised in Table 3. *Bl*Arap42B is homo-trimer of a 701 residues subunit, with three subunits related by a 3-fold axis (Figure 2*A*). Each subunit consists of three domains. Domain A is a $(\beta/\alpha)_8$ barrel containing the catalytic residues. Domain B is a structural domain that packs onto Domain A of the adjacent monomer. Domain C, whose role is unknown, adopts an anti-parallel β-sandwich fold. The first seven residues (Met1–Asp7) and the region Pro663–Thr672 in all six chains, as well as a variable number of residues in the C-terminus of each chain (Gly695–Asn701), were disordered and not included in the final model.

Bl*Arap42B active-site architecture and ligand docking*–Similar to structures of GH42 β-galactosidases, the active sites and the vicinity encompassing the proposed subsites +1 and +2 are formed at the interface of two adjacent monomers in the *Bl*Arap42B trimer (18). It was possible to dock an energetically preferred conformation of paeonolide, the preferred substrate for *Bl*Arap42B, into the active site with an estimated affinity of −8.8 kcal mol^{-1} (Figure 2*B*). The docked paeonolide molecule makes an intramolecular hydrogen bond between the acetyl oxygen of the aglycon (Figure 1*A*) and C2-OH of the Glc*p*. The Ara*p* ring binds at subsite −1, with distances of 3.3 Å between the catalytic nucleophile Glu311 and the anomeric carbon and 2.9 Å between the catalytic acid/base Glu151 and the glycosidic oxygen, respectively. The Glc*p* ring bound at subsite +1, is stacking onto Trp252, and C4-OH is hydrogen bonded with Asp249. Furthermore, the aromatic ring of the aglycone and Phe421 from the neighboring monomer are almost parallel, making an aromatic stacking at this position, indicative of the presence of a subsite +2.

Comparison with the Bl*Gal42A β-galactosidase structure*–The structure of *Bl*Arap42B was compared to that of the other GH42 enzyme from the same organism (*Bl*Gal42A) that has specificity towards β(1,6)/β(1,3)-galactoside linkages and is one of the best characterized enzymes (both biochemically and structurally) among GH42 β-galactosidases (13). Figure 3A shows superimposition of the structures of *Bl*Arap42B and *Bl*Gal42A in complex with galactose (PDB: 4UNI). From the structural comparison, Glu151 and Glu311 of *Bl*Arap42B were inferred as the acid/base catalyst and the catalytic nucleophile, respectively. Although *Bl*Arap42B and *Bl*Gal42A only share 25% sequence identity their overall structures are very similar as reflected by the root mean square deviation (RMSD) for C_α atoms of 1.75 Å when chain A (689/695 residues) of *Bl*Gal42A is superimposed on chain A (682/701 residues) of *Bl*Arap42B.

Remarkably, three residues at the substrate binding site are either variant or spatially differently located in *Bl*Arap42B as compared to *Bl*Gal42A (Figure 3*A*). Trp345 is phenylalanine in *Bl*Gal42A (Phe362) and the aromatic side chain is involved in substrate recognition in subsite −1 by forming a hydrophobic platform for the C4 side of the arabino- or galactopyranoside. The change of a histidine in *Bl*Gal42A (His375) to Trp358 restricts the space in subsite −1 around the C6-O6 hydroxymethyl group of galactoside, which is not present in arabinopyranoside (Figure 3*B*). Moreover, Trp332 of *Bl*Gal42A makes a hydrogen

bond to the C6-OH of galactose, and is also proposed to mediate aromatic stacking to substrate at subsite +1 in GH42 β-galactosidases. In contrast, the conformation of the loop containing Trp332 is different in *Bl*Arap42B and locates a tryptophan at the corresponding position in the sequence (Trp320, not shown in Figure 3*A*), distantly from the active site in space. Instead, a differently located Trp252, which is not conserved in previously characterized GH42 enzymes, is taking the role of stacking platform at subsite +1 in *Bl*Arap42B. Additionally, Met268 in *Bl*Gal42A replaces Asp249 that in the docked structure of *Bl*Ara42B forms a hydrogen bond to the Glc*p* C4-OH in paeonolide.

Mutational analysis–From the structural comparison with *Bl*Gal42A, two non-conserved residues (Trp345 and Trp358) are identified in subsite –1. To examine the role of these residues in relation to substrate specificity, these single residues were replaced using site-directed mutagenesis with those found in the β-galactosidase (Phe or His). Table 1 shows specific activities of the single (W345F and W358H) and double (W345F/W358H) mutants towards *p*NP-substrates, the single mutants having more than 10-fold reduced activity on *p*NP-α-L-Ara*p* and the double mutant being almost inactive. W358H was slightly active on *p*NP-β-D-Fuc*p*, but none of the mutants showed detectable activity on *p*NP-β-D-Gal*p* or any of the other *p*NP-substrates. Clearly, the substrate specificity of the GH42 α-L-arabinopyranosidase was not changed to β-galactosidase by these simple mutations.

Phylogeny and active site motifs of GH42–A phylogenetic and active site sequence motif analysis of GH42 α-L-arabinopyranosidases (subfamily A) and β-galactosidases (subfamily G), divides GH42 into two subfamilies, respectively (Figure 4). *Bl*Arap42B assigns into a distinct uncharacterized clade, group one (G1) in the phylogenetic tree. Based on sequence analysis, it is evident that the two tryptophan residues (Trp345 and Trp358) in *Bl*Arap42B are conserved in G1 and subfamily A, and comprise a unique Trp-X_{12}-Trp sequence motif through the active site as compared to subfamily G that includes G2–G4. All of the G2–G4 groups of subfamily G contain at least one characterized β-galactosidase that has a Phe-X_{12}-His motif. Trp252 of *Bl*Arap42B in subsite +1 is conserved in G1 but not in G2–G4 (data not shown).

DISCUSSION

Bl*Arap42B is a novel GH42 α-L-arabinopyranosidase*–The GH42 family has been assumed very homogenous since all hitherto characterized members exhibited β-galactosidase activity despite representing broad diverse taxa. This further indicated active site conservation within the family, where evolution of the molecular machinery was driven towards subtle changes in the plus subsites for fine-tuning subspecificities, *e.g.* for β(1,6), β(1,4), or β(1,3)-linked galactosides or *N*-acetyl-glucosides (22–24).

While GH42 *Bl*Arap42B from *B. animalis* subsp. *lactis* Bl-04 was unable to hydrolyze synthetic β-galactosides and those from plants, mammalian milk and host mucin, so far demonstrated as substrates for GH42 β-galactosidases, it was able to hydrolyze, the different yet structurally similar, α-L-arabinopyranosides (Tables 1, 2 and Figure 1). *Bl*Arap42B effectively released α-L-arabinopyranose from paeonolide and ginsenoside Rb2, both containing an aglycone conjugated to α-L-Ara*p*-(1,6)-β-D-Glc*p*, but was unable to hydrolyze quercetin 3-*O*-α-L-Ara*p*. The subsite +1 is apparently specific for the β(1,6)-linked glucoside with Trp252 as a sugar-stacking residue (Figure 3A), and aromatic residues surrounding this site (Trp358, Phe354, and Phe421) probably make steric clash with the large flavonoid group of quercetin 3-*O*-α-L-Ara*p*. Thus, *Bl*Arap42B have the capabilities to discriminate substrates beyond subsite −1 and demonstrate the potential existence of different α-L-arabinopyranoside sub-specificities. Previously characterized bifidobacterial GH42 β-galactosidases tested here, did not cleave paeonolide, although they were able to hydrolyze *p*NP-α-L-Ara*p* (Figure 1*B*). *Bl*Arap42B is clearly a GH42 α-L-arabinopyranosidase, distinctly different from previously characterized GH42 enzymes. The k_{cat} (240 s^{-1}) and K_m (0.074 mM) values of *Bl*Arap42B towards its preferred substrate, paeonolide, are in the same range as those of previously characterized GH42 enzymes towards their preferred substrates (13, 24).

Structural specificity determinants in GH42 α-L-arabinopyranosidases–The residues creating the spatial and chemical environment at subsite −1 are invariant in characterized GH42 β-galactosidases (13, 18, 22, 26, 27). However, the change from a histidine residue in classical GH42 β-galactosidases to Trp358 in *Bl*Arap42B limits the space at subsite −1, which would cause clashing with C6 of galactose in accordance with *Bl*Arap42B being unable to hydrolyze β-galactosides (Figure 3*B*). A similar structural change is observed between GH27 α-D-galactosidases and β-L-arabinopyranosidases, the only other similar described activity. In GH27 β-L-arabinopyranosidases, a mutagenesis study revealed that the single substitution from an aspartic acid to the slightly larger glutamic acid in the catalytic pocket around the C-6 of galactose as compared to GH27 α-D-galactosidases is critical for modulating the enzyme activity (28). However, a similar attempt to change the specificity of *Bl*Arap42B from α-L-arabinopyranosidase to β-galactosidase through the single and double mutations of Trp345 and Trp358 failed (Table 1). This may suggest that the active site of this enzyme is optimized to α-L-arabinopyranoside through accumulation of other substitutions around subsite −1 during the molecular evolution after it diverged from the β-galactosidases.

Notably, a major change is observed in the positioning of the subsite +1 stacking platform and potential recognition of hydroxyl groups between *Bl*Arap42B and *Bl*Gal42A (Figure 3*A*). Superimposing the docked paeonolide from *Bl*Arap42B into *Bl*Gal42A shows the phenyl aglycone clashes with the backbone of Trp332, which is invariant and similarly spatially located in structurally characterized GH42 enzymes (13, 18, 22, 26, 27), and Phe226 of the neighboring monomer in *Bl*Gal42A, which can explain its lack of activity for paeonolide (Figure 3*C*). The present findings indicate that GH42 α-L-arabinopyranosidases have evolved from a common scaffold to target specific α-L-arabinopyranosides present in plant oligosaccharides.

Bl*Arap42B defines a novel GH42 α-L-arabinopyranosidase subfamily*–Applying a phylogenetic and sequence analysis approach links the α-L-arabinopyranoside specificity of *Bl*Arap42B with a unique Trp-X_{12}-Trp sequence motif at the active site, as opposed to Phe-X_{12}-His in classical bifidobacterial GH42 β-galactosidases (Figure 4). The Trp-X_{12}-Trp sequence motif is present in enzymes with a GH42 catalytic domain from more than 150 different bacterial species and subspecies of various phyla, and uncovers a novel α-L-arabinopyranosidase subfamily (subfamily A) distinct from GH42 β-galactosidases (subfamily G). As a confirmation that the subfamily division reflects specificity, we have shown that *Ri*Arap42B, a homologue and subfamily A member from a human gut bacterium, *Roseburia intestinalis* M50/1 (29), shows the same activity profile on *p*NP-derived substrates, hydrolyzing *p*NP-α-L-Ara*p*, but not *p*NP-β-D-Gal*p* (Table 1), supporting the exclusive activity towards α-L-arabinopyranoside of this subfamily. Noticeably, BgaA from *Clostridium cellulovorans* (30), belonging to subfamily G, contains a Phe-X_{12}-Trp-mixed sequence motif and is the only α-L-arabinopyranosidase currently assigned to GH42 in the CAZy database. *Clostridium cellulovorans* BgaA, however, hydrolyzes *p*NP-β-D-Gal*p* with about 10% activity of *p*NP-α-L-Ara*p* (30) as opposed to *Bl*Arap42B that shows <0.15% activity of *p*NP-α-L-Ara*p* on *p*NP-β-D-Gal*p*.

Subspecificities within the α-L-arabinopyranosidase subfamily A–Distinct groups clearly exist within subfamily A (Figure 4), with members of different groups organized in different gene landscapes. The gene encoding *Bl*Arap42B is located adjacent to a gene encoding a putative GH30 enzyme, with no close characterized homologues based on sequence similarity, as well as in the near vicinity of a putative GH3 enzyme (*Bl*GH3). The closest characterized homologue to *Bl*GH3 is the GH3 Apy-H1 from *B. longum* H-1, showing 61% sequence identity and 74% similarity, and dual α-L-arabinopyranosidase / β-galactosidase specificity, releasing α-L-arabinopyranoside from ginsenoside Rb2 found in *Panax ginseng* (31). Interestingly, *Bl*Arap42B is also able to release α-L-arabinopyranoside from ginsenoside Rb2 with high efficiency. This gene cluster in *B. animalis* subsp. *lactis* Bl-04 likely targets α-L-arabinopyranosides found in plants of especially Asian origin (17, 32). Similarly, a GH3

member is encoded adjacent to the gene of *Ri*Arap42B in *R. intestinalis* M50/1, but it clusters with characterized xylan β(1,4) xylosidases based on a phylogenetic analysis of this family (data not shown).

The genes surrounding the gene encoding a putative GH42 α-L-arabinopyranosidase from *Bacteriodes uniformis* ATCC 8492 (*Bu*42B) of subfamily A (Figure 4) contains all glycoside hydrolase-encoding genes necessary for complete utilization of xyloglucan oligosaccharides (33), suggesting that α-L-arabinopyranosides exist in xyloglucan, but this has yet to be identified. Generally, α-L-arabinopyranoside is sparsely reported in plant oligosaccharides, which is in sharp contrast to the evident appearance of *Bl*Arap42B homologues in different gene landscapes found throughout the bacterial kingdom. GH42 α-L-arabinopyranosidases therefore may prove useful in the identification of α-L-arabinopyranosides in plant glycans, oligosaccharides and glycoconjugates.

α-L-Arabinopyranosidase activity in the human gut niche–Recently, the human gut microbe *Bacteriodes thetaiotaomicron* was shown to be capable to completely depolymerise the highly complex pectic Rhamnogalacturonan II, containing buried α-L-Ara*p*-(1,4)-β-D-Gal*p*, through the novel BT0983 GH2 α-L-arabinopyranosidase (34). Additionally, the human gut microbe *B. breve* K-110, isolated from healthy adults, can utilize ginsenoside Rb2 (35). However, the internal amino acid sequence (VIYLTDA) of the purified α-L-arabinopyranosidase from *B. breve* K-110 match with an AAA^+ family ATPase but not with any GHs in the genome sequences, suggesting that the sequence was derived from a contaminated protein. Since ginsenoside Rb2 was effectively hydrolyzed by *Bl*Arap42B, *B. animalis* subsp. *lactis* Bl-04 is potentially conferred with the ability to utilize this compound. The present study reports the enzymology and structure of a GH42 α-L-arabinopyranosidase from the probiotic *B. animalis* subsp. *lactis* Bl-04, uncovering the existence of a distinct GH42 subfamily, with novel specificity towards α-L-arabinopyranosides and identifying a unique associated active site sequence motif. The molecular insights together with a deep phylogenetic analysis identify key structural elements discriminating the GH42 α-L-arabinopyranosidases from the classical GH42 β-galactosidases. The phylogenetic analysis uncovered the existence of unexpectedly large number (>150) of potential GH42 α-L-arabinopyranosidases in bacterial genomes. It is noteworthy that the plant glycosides effectively hydrolyzed by *Bl*Arap42B, paeonolide and ginsenoside Rb2, are contained in traditional herbal medicines in East Asian countries, such as Cortex Moutan (*Mu Dan Pi* in Chinese and *Botanpi* in Japanese) and Asian Ginseng, suggesting that other edible root vegetables may also contain certain amounts of glycosides with an α-L-arabinopyranoside. Altogether, these results indicate that α-L-arabinopyranoside metabolism evolved in different ecological niches and probably is widespread in important probiotic members of the human gut microbiota.

EXPERIMENTAL PROCEDURES

Substrates–*p*NP-β-D-galactopyranoside (*p*NP-β-D-Gal*p*), *p*NP-β-D-glucopyranoside (*p*NP-β-D-Glc*p*), *p*NP-α-L-arabinofuranoside (*p*NP-α-L-Ara*f*), *p*NP-α-L-arabinopyranoside (*p*NP-α-L-Ara*p*), *p*NP-β-D-fucopyranoside (*p*NP-β-D-Fuc*p*), and *p*NP-β-D-xylopyranoside (*p*NP-β-D-Xyl*p*) were purchased from Sigma-Aldrich (St. Louis, MO, USA). 4-β-galactobiose, 6-β-galactobiose, LacNAc were purchased from Dextra Laboratories (Readings, UK). Lacto-*N*-biose I (LNB) and galacto-*N*-biose (GNB) were synthesised as described previously (36, 37). Lacto-*N*-tetraose (LNT) and lacto-*N*-*neo*tetraose (LN*n*T) were purchased from Elicityl (Crolles, France). 2-Acetyl-5-methoxyphenyl α-L-Ara*p*-(1,6)-β-D-Glc*p* (paeonolide), ginsenoside Rb2, and quercetin 3-*O*-α-L-Ara*p* was purchased from Tokyo Chemical Industry (Tokyo, Japan), Carbosynth (Berkshire, UK), and eNovation Chemicals LLC (Bridgewater, NJ, USA), respectively.

Expression and Purification–The gene encoding a putative GH42 enzyme (locus tag: *balac_*0053, GenBank accession number ACS45449.1) was amplified from *Bifidobacterium animalis* subsp. *lactis* Bl-04 genomic DNA by PCR using High-fidelity DNA polymerase (Fermentas, St. Leon-Rot, Germany) with the following primers: 5'-CTA GCT AGC GCC CGC GCA TAC ACC-3'; reverse: 5'-ATA AGA ATG CGG CCG CAT TCA

TTG TGG GTT GC-3'. Amplified DNA was cloned into pET21(a)+ (Novagen, Darmstadt, Germany) using the NheI and NotI restriction sites (underlined) resulting in *Bl*Arap42B with a C-terminal His-tag. *Escherichia coli* BL21 (DE3) Δ*lacZ*(β-Gal) containing the above expression plasmid was grown at 20 °C in 1 L Luria-Bertani medium containing 50 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol to an OD_{600} of 0.5, and expression was induced by 100 μM isopropyl-β-thiogalactopyranoside for 36 h at 18 °C. The recombinant enzyme (*Bl*Arap42B) was purified using a 5 mL HisTrap HP column (GE Healthcare) with a linear gradient 2.5–100% 400 mM imidazole (20 mM HEPES buffer, 0.5 M NaCl, pH 7.5). The His-purified fractions were pooled and purified by gel filtration (HiLoad 26/60 Superdex G200; GE Healthcare) in 10 mM MES, 150 mM NaCl (pH 6.5) eluted by 1.2 column volume of this buffer at a flow rate of 1 mL min^{-1}. The purity of the enzyme was analysed by SDS-PAGE. The enzyme concentration was determined spectrophotometrically (ε_{280} = 152,890 M^{-1} cm^{-1}). The W345F and W358H, and double W345F/W358H mutants were constructed using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, California, USA), and the recombinant proteins were prepared and purified as described for the wild-type. The gene with locus tag Roi_37780 from *Roseburia intestinalis* M50/1 (GenBank accession number CBL10553.1) encoding a homologue of *Bl*Arap42B was synthesised with a C-terminal His-tag and cloned into pET21(a)+ (GenScript, USA), expressed and the resulting recombinant protein (*Ri*Arap42B), was purified as described above for *Bl*Arap42B. Bga42A, Bga42B, and Bga42C from *B. longum* subsp. *infantis* ATCC 15697, and *Bl*Gal42A from *B. animalis* subsp. *lactis* Bl-04 were prepared as previously described (13, 24).

Biochemical Characterization–Activity was assayed towards *p*NP-β-D-Gal*p*, *p*NP-α-L-Ara*p*, *p*NP-β-D-Fuc*p*, *p*NP-α-L-Ara*f*, *p*NP-β-D-Glc*p*, and *p*NP-β-D-Xyl*p* (final conc. 5 mM) at 37 °C in 40 mM sodium citrate, 0.005% Triton X-100, pH 6.5 (50 μL) by addition of enzyme (2–2,000 nM) and stopping the reaction after 10 min by 1 M Na_2CO_3 (200 μL) for *Bl*Arap42B wild-type and mutants, and *Ri*Ara42B. The amount of released *p*NP was measured spectrophotometrically at A_{410} using *p*NP (0–2 mM) as standard. One unit of activity was defined as the amount of enzyme that released 1 μmol *p*NP min^{-1}. The kinetic parameters k_{cat} and K_m were determined from initial rates of *p*NP-α-L-Ara*p* (0.5–7.5 mM) hydrolysis in the above buffer by non-linear regression fit of the Michaelis-Menten model: $v = k_{cat} \times [E] \times [S] / (K_m + [S])$ to data from triplicate experiments (GraphPad Prism 6, La Jolla, USA). Activity of *Bl*Arap42B as well as Bga42A, Bga42B, and Bga42C from *B. longum* subsp. *infantis* ATCC 15697, and *Bl*Gal42A from *B. animalis* subsp. *lactis* Bl-04 (75–125 nM) were measured towards 4 mM *p*NP-α-L-Ara*p* and 4 mM paeonolide in 50 mM sodium phosphate pH 6.5 at 30°C for 30 min and 24 h and monitored by TLC. HPAEC-PAD experiments were carried out using a Dionex ICS-3000 HPLC system (Chromeleon software version 6.80, Dionex) with a Dionex Carbopac PA20 column. The analytes were separated using 40 mM NaOH in isocratic mode at 0.5 ml min^{-1}. Initial hydrolysis activities were measured using 20 nM *Bl*Arap42B at 30 °C in 40 mM sodium citrate (pH 6.5) towards paeonolide, ginsenoside Rb2, and quercetin 3-*O*-α-L-Ara*p* (100 μM) in 19.2 – 65.2 min assays. The kinetic parameters k_{cat} and K_m were determined from initial rates of paeonolide, and ginsenoside Rb2 (7.812 – 1000 μM) using 40 pM *Bl*Arap42B at 30 °C in the above buffer at 4 time points by integrating the area of the peaks corresponding to released L-arabinose.

Crystallization, Data Collection, and Structure Determination of BlArap42B–*Bl*Arap42B was concentrated to 12.0 mg mL^{-1} in 10 mM MES pH 6.5, 150 mM NaCl and screened for initial crystallization conditions at 20 °C with the JCSG core I–IV screens (Qiagen). Crystals were observed in the JCSG core IV screen (1.5 M ammonium sulfate, 12% glycerol (v/v), 0.1 M Tris pH 8.5). The final crystallization condition was 12% glycerol, 1.6 M ammonium sulfate, 0.1 M MES buffer pH 6.5 obtained from optimization in 24-well VDX trays (Hampton Research) in sitting drops containing 1 μL protein stock and 1 μL reservoir at 20 °C. Glycerol (20% final conc.) was added as cryoprotectant before harvesting.

A data set to 2.0 Å resolution was obtained at the BL5A beamline, Photon Factory, Tsukuba,

Japan. Processing and scaling of the data was done with HKL2000 (38). The space group was determined to be $P4_12_12$ with six molecules in the asymmetric unit. Molecular Replacement was performed using Balbes (39) with a chain of the homotrimer of the PDB entry code 1KWG (18) as the initial search model. ARP/wARP (40) was used to partially build the structure, and refinement was completed using Coot (41) and Refmac5 (42).

The quality of the structures, including Ramachandran statistics, was verified by MolProbity (43), and PyMOL v1.8.5 (Schrödinger, LLC, New York) was used for structural analysis and rendering of molecular graphics.

Ligand preparation and docking–Paeonolide was prepared from the PubChem project (CID: 92043525) and PCModel v 9.20 (Serena Software) using energy minimisation with MMX force field. The energy minimised molecule was docked into the active site of *Bl*Arap42B using AutoDock Vina 1.1.2 (44) with the grid box (16 Å × 16 Å × 16 Å) centred on the scissile glycosidic bond oxygen. The ligand structure was docked with flexible torsion angles, whereas the protein structure was fixed.

Bioinformatics analysis–The amino acid sequences of bifidobacterial GH42 members were extracted from the CAZy database (7) using CAZy Tools (A. H. Viborg, http://research.ahv.dk/cazy) and aligned using MUSCLE v3.8 with default settings (45). The active site sequence motif (WHWHSIHNSFETYW) for group one (G1) and (FQWRQSRGGAEKFH) for group two, three, and four (G2, G3, and G4) were queried against the NCBI nr-database using BLASTP. Non-identical sequences were extracted and the redundancy was decreased by USEARCH (46) to yield approximately 200 sequences from each of the two sequence motifs, which were merged with the bifidobacterial entries and previously characterized GH42 enzymes. Sequence motif logos for the subfamilies were rendered by WebLogo 3 (47). ClustalW2 was used to render a phylogenetic tree with default settings, which was visualised in Dendroscope 3.0 (48).

Acknowledgements: AHV was an International Research Fellow of the Japan Society for the Promotion of Science. We thank Dr. M. Hidaka and the staff of the Photon Factory and SPring-8 for the X-ray data collection. We thank Prof. Harry Brumer for laboratory facilities, as well as Dr. Shaheen Shojania and Mr. Nicholas McGregor for their laboratory assistance.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: AHV, MAH, and BS conceived the study. AHV conducted most of the experiments and analysed the results. AHV wrote the manuscript with contributions from SF. TK conducted TLC experiments on the hydrolysis of paeonolide. TA and SF assisted in X-ray data collection, structure determination and analysis. SF performed molecular replacement and ligand docking. LLL assisted in preliminary crystallisation trials, structural analysis and verification. TK, MAH, MK and BS assisted in experimental design and analysis of the data. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012) Host-gut microbiota metabolic interactions. *Science*. **336**, 1262–7
2. Koropatkin, N., Cameron, E., and Martens, E. (2012) How glycan metabolism shapes the human gut microbiota. *Nat. Rev.* **10**, 323–35
3. Peacock, K. S., Ruhaak, L. R., Tsui, M. K., Mills, D. A., and Lebrilla, C. B. (2013) Isomer-specific consumption of galactooligosaccharides by bifidobacterial species. *J. Agric. Food Chem.* **61**, 12612–9
4. Han, N. S., Kim, T.-J., Park, Y.-C., Kim, J., and Seo, J.-H. (2012) Biotechnological production of human milk oligosaccharides. *Biotechnol. Adv.* **30**, 1268–78
5. Ridley, B. L., O'Neill, M. A., and Mohnen, D. (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*. **57**, 929–67
6. Vincken, J., Schols, H. A., Oomen, R. J. F. J., Mccann, M. C., Ulvskov, P., Voragen, A. G. J., and Visser, R. G. F. (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. *Plant Physiol.* **132**, 1781–9
7. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-5
8. Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B., German, J. B., Price, N. P., Richardson, P. M., and Mills, D. A. (2008) The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18964–9
9. Viborg, A. H., Katayama, T., Abou Hachem, M., Andersen, M. C., Nishimoto, M., Clausen, M. H., Urashima, T., Svensson, B., and Kitaoka, M. (2014) Distinct substrate specificities of three glycoside hydrolase family 42 β-galactosidases from *Bifidobacterium longum* subsp. *infantis* ATCC 15697. *Glycobiology*. **24**, 208–16
10. James, K., Motherway, M. O., Bottacini, F., and van Sinderen, D. (2016) *Bifidobacterium breve* UCC2003 metabolises the human milk oligosaccharides lacto-*N*-tetraose and lacto-*N*-neo-tetraose through overlapping, yet distinct pathways. *Sci. Rep.* **6**, 38560
11. Motherway, M. O., Fitzgerald, G. F., and van Sinderen, D. (2011) Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb. Biotechnol.* **4**, 403–16
12. Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sánchez, B., Bidossi, A., Ferrarini, A., Giubellini, V., Delledonne, M., Henrissat, B., Coutinho, P., Oggioni, M., Fitzgerald, G. F., Mills, D., Margolles, A., Kelly, D., van Sinderen, D., and Ventura, M. (2010) Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19514–9
13. Viborg, A. H., Fredslund, F., Katayama, T., Nielsen, S. K., Svensson, B., Kitaoka, M., Lo Leggio, L., and Abou Hachem, M. (2014) A β1-6/β1-3 galactosidase from *Bifidobacterium animalis* subsp. *lactis* Bl-04 gives insight into sub-specificities of β-galactoside catabolism within *Bifidobacterium*. *Mol. Microbiol.* **94**, 1024–40
14. Barrangou, R., Briczinski, E. P., Traeger, L. L., Loquasto, J. R., Richards, M., Horvath, P., Coûté-Monvoisin, A.-C., Leyer, G., Rendulic, S., Steele, J. L., Broadbent, J. R., Oberg, T., Dudley, E. G., Schuster, S., Romero, D. A., and Roberts, R. F. (2009) Comparison of the complete genome sequences of *Bifidobacterium animalis* subsp. *lactis* DSM 10140 and Bl-04. *J. Bacteriol.* **191**, 4144–51
15. Andersen, J. M., Barrangou, R., Abou Hachem, M., Lahtinen, S. J., Goh, Y. J., Svensson, B., and Klaenhammer, T. R. (2013) Transcriptional analysis of oligosaccharide utilization by *Bifidobacterium lactis* Bl-04. *BMC Genomics*. **14**, 312
16. Viborg, A. H. (2015) Diversity in β-galactosidase specificities within *Bifidobacterium*: towards an understanding of β-galactoside metabolism in the gut niche. *Trends Glycosci. Glycotechnol.* **27**,

E9–E12
17. Ding, L., Zuo, Q., Li, D., Feng, X., Gao, X., Zhao, F., and Qiu, F. (2017) A new phenone from the roots of *Paeonia suffruticosa* Andrews. *Nat. Prod. Res.* **31**, 253–60
18. Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakagi, T. (2002) Trimeric crystal structure of the glycoside hydrolase family 42 β-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Mol. Biol.* **322**, 79–91
19. Oura, H., Hiai, S., Odaka, Y., and Yokazawa, T. (1975) Studies on the Biochemical Ation of *Ginseng* Saponin. *J. Biochem.* **77**, 1057–65
20. Fraisse, D., Heitz, A., Carnat, A., Carnat, A. P., and Lamaison, J. L. (2000) Quercetin 3-arabinopyranoside, a major flavonoid compound from *Alchemilla xanthochlora*. *Fitoterapia*. **71**, 463–4
21. Metwally, A. M., Omar, A. A., Harraz, F. M., and El Sohafy, S. M. (2010) Phytochemical investigation and antimicrobial activity of *Psidium guajava* L. leaves. *Pharmacogn Mag*. **6**, 212–8
22. Godoy, A. S., Camilo, C. M., Kadowaki, M. A., Muniz, H. dos S., Espirito Santo, M., Murakami, M. T., Nascimento, A. S., and Polikarpov, I. (2016) Crystal structure of β1→6-galactosidase from Bifidobacterium bifidum S17: trimeric architecture, molecular determinants of the enzymatic activity and its inhibition by α-galactose. *FEBS J.* **283**, 4097–112
23. Tabachnikov, O., and Shoham, Y. (2013) Functional characterization of the galactan utilization system of *Geobacillus stearothermophilus*. *FEBS J.* **280**, 950–64
24. Yoshida, E., Sakurama, H., Kiyohara, M., Nakajima, M., Kitaoka, M., Ashida, H., Hirose, J., Katayama, T., Yamamoto, K., and Kumagai, H. (2012) *Bifidobacterium longum* subsp. *infantis* uses two different β-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology*. **22**, 361–8
25. Van Laere, K. M. J., Abee, T., Schols, H. A., Beldman, G., and Voragen, A. G. J. (2000) Characterization of a novel β-galactosidase from *Bifidobacterium adolescentis* DSM 20083 active towards transgalactooligosaccharides. *Appl. Environ. Microbiol.* **66**, 1379–84
26. Maksimainen, M., Paavilainen, S., Hakulinen, N., and Rouvinen, J. (2012) Structural analysis, enzymatic characterization, and catalytic mechanisms of β-galactosidase from *Bacillus circulans* sp. *alkalophilus*. *FEBS J.* **279**, 1788–98
27. Solomon, H. V., Tabachnikov, O., Lansky, S., Salama, R., Feinberg, H., Shoham, Y., and Shoham, G. (2015) Structure-function relationships in Gan42B, an intracellular GH42 β-galactosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 2433–2448
28. Ichinose, H., Fujimoto, Z., Honda, M., Harazono, K., Nishimoto, Y., Uzura, A., and Kaneko, S. (2009) A β-L-arabinopyranosidase from *Streptomyces avermitilis* is a novel member of glycoside hydrolase family 27. *J. Biol. Chem.* **284**, 25097–105
29. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S., and Flint, H. J. (2002) *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620
30. Kosugi, A., Murashima, K., and Doi, R. (2002) Characterization of two noncellulosomal subunits, ArfA and BgaA, from *Clostridium cellulovorans* that cooperate with the cellulosome in plant cell wall degradation. *J. Bacteriol.* **184**, 6859–65
31. Lee, J. H., Hyun, Y.-J., and Kim, D.-H. (2011) Cloning and characterization of α-L-arabinofuranosidase and bifunctional α-L-arabinopyranosidase/β-D-galactopyranosidase from *Bifidobacterium longum* H-1. *J. Appl. Microbiol.* **111**, 1097–107
32. Lee, D. Y., Jeong, Y. T., Jeong, S. C., Lee, M. K., Min, J. W., Lee, J. W., Kim, G. S., Lee, S. E., Ahn, Y. S., Kang, H. C., and Kim, J. H. (2015) Melanin biosynthesis inhibition effects of ginsenoside Rb2 isolated from panax ginseng berry. *J. Microbiol. Biotechnol.* **25**, 2011–2015
33. Larsbrink, J., Rogers, T., and Hemsworth, G. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature*. **506**, 498–502
34. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y.,

O'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M.-C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*. 10.1038/nature21725

35. Shin, H., Park, S., Sung, J. H., and Kim, D. (2003) Purification and characterization of α-L-arabinopyranosidase and α-L-arabinofuranosidase from *Bifidobacterium breve* K-110, a human intestinal anaerobic bacterium metabolizing ginsenoside Rb2 and Rc. *Appl. Environ. Microbiol.* **69**, 7116–23
36. Nishimoto, M., and Kitaoka, M. (2007) Practical preparation of lacto-*N*-biose I, a candidate for the bifidus factor in human milk. *Biosci. Biotechnol. Biochem.* **71**, 2101–4
37. Nishimoto, M., and Kitaoka, M. (2009) One-pot enzymatic production of β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactose (galacto-*N*-biose) from sucrose and 2-acetamido-2-deoxy-D-galactose (*N*-acetylgalactosamine). *Carbohydr. Res.* **344**, 2573–6
38. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–26
39. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2007) *BALBES*: A molecular-replacement pipeline. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 125–32
40. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–63
41. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *Coot*. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501
42. Nicholls, R. A., Long, F., and Murshudov, G. N. (2012) Low-resolution refinement tools in *REFMAC5* research papers. *Acta Crystallogr. D. Biol. Crystallogr.* **D68**, 404–17
43. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21
44. Trott, O., and Olson, A. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–61
45. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–7
46. Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. **26**, 2460–1
47. Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004) WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–90
48. Huson, D. H., and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–7

FOOTNOTES

This work was supported by Grant-in-Aid for JSPS Research Fellows (AHV), and in part by JSPS KAKENHI Grant Numbers 15H02443 and 26660083 (to SF) and the Platform for Drug Discovery, Informatics, and Structural Life Science funded by MEXT. Technical University of Denmark supported initial work through a PhD fellowship (to AHV).

The atomic coordinates and structure factors (code 5XB7) have been deposited in the Protein Data Bank (http://wwpdb.org/).

The abbreviations used are as follows: CAZy, Carbohydrate Active enZymes; GH, glycoside hydrolase; GNB, galacto-*N*-biose; GOS, galacto-oligosaccharide; HPAEC-PAD, high-performance anion-exchange chromatography with pulsed amperometric detection; LNB, lacto-*N*-biose I; LN*n*T, lacto-*N*-*neo*tetraose; LNT, lacto-*N*-tetraose; *p*NP, *p*-nitrophenyl; RMSD, root mean square deviation.

FIGURE LEGENDS

FIGURE 1. (A) Chemical structures of substrates, (B) TLC analysis on the hydrolysis of *p*NP-α-L-Ara*p* and paeonolide by *Bl*Arap42B and other GH42 enzymes, and (C) HPAEC-PAD analysis on the hydrolysis of plant glycosides containing a non-reducing end α-L-arabinopyranoside by *Bl*Arap42B.
(A) Structures of *p*NP-α-L-Ara*p*, *p*NP-β-D-Fuc*p,* *p*NP-β-D-Gal*p*, and paeonolide, ginsenoside Rb2, and quercetin 3-*O*-α-L-Ara*p* are shown. *p*NP-α-L-Ara*p* and *p*NP-β-D-Fuc*p* differ from *p*NP-β-D-Gal*p* by lacking C6 and C6-OH, respectively. Non-reducing end α-L-arabinopyranose is found in 2-acetyl-5-methoxyphenyl-α-L-Ara*p*-(1,6)-β-D-Glc*p* (paeonolide) from *Paeonia*, ginsenoside Rb2 from *Panax ginseng* and quercetin 3-*O*-α-L-Ara*p* from *Psidium guajava* Linn. **(B)** Substrate activity screening using TLC of *Bl*Arap42B (4) and previously characterized bifidobacterial GH42 β-galactosidases, Bga42A (1), Bga42B (2), and Bga42C (3) from *B. longum* subsp. *infantis* ATCC 15697 as well as *Bl*Gal42A (5) from *B. animalis* subsp. *lactis* Bl-04. (0) is control without enzyme. *Bl*Arap42B (4) displayed significant activity towards paeonolide by 30 min incubation, with complete hydrolysis overnight. Bga42A, Bga42B, Bga42C, and *Bl*Gal42A did not show significant activity towards paeonolide after 24 h incubation, despite their ability to hydrolyze *p*NP-α-L-Ara*p*. **(C)** HPAEC-PAD chromatograms showing activity towards paeonolide and ginsenoside Rb2. Paeonoside is the de-arabinosylated product of paeonolide. Activity was not detected towards quercetin 3-*O*-α-L-Ara*p*. Peaks corresponding to the substrate ginsenoside Rb2 and its de-arabinosylated product (= ginsenoside Rd) could not be detected in the elution range under the conditions employed (see Experimental Procedures).

FIGURE 2. Overall structure and docking of paeonolide into *Bl*Arap42B.
(A) Top view along the axis of the *Bl*Arap42B trimer with docked paeonolide shown as spheres in one of the three active sites. **(B)** Active site residues of *Bl*Arap42B shown with docked paeonolide. The Ara*p* moiety binds at subsite −1 with distances of 3.3 Å between the catalytic nucleophile Glu311 and the anomeric carbon, and 2.9 Å between the catalytic acid/base Glu151 and the glycosidic oxygen. Glc*p* participates in an aromatic stacking interaction with the proposed subsite +1 platform Trp252 and C4-OH hydrogen bonds to Asp249. Paeonolide in this conformation can form an intramolecular hydrogen bond between its acetyl oxygen and C2-OH of the Glc*p*. The neighboring monomer participates in determining the substrate specificity by Phe421 engaged in aromatic stacking with the phenyl group of paeonolide at a possible subsite +2.

FIGURE 3. Structural comparison of *Bl*Arap42B α-L-arabinopyranosidase and *Bl*Gal42A β-galactosidase.
(A) Stereo view of the comparison of the active site with subsites −1 through +2 of *Bl*Arap42B (green) and the *Bl*Gal42A β-galactosidase (white and grey) in complex with galactose from *B. animalis* subsp. *lactis* Bl-04 (PDB: 4UNI). Trp345 and Trp358 occupy the active site environment in *Bl*Arap42B, as compared to the smaller side chains of Phe362 and His375 in *Bl*Gal42A. The subsite +1 aromatic stacking platform is composed of Trp residues from different loops (Trp252 in *Bl*Arap42B and Trp332 in *Bl*Gal42A) and are spatially differently located in the two enzymes. **(B)** The residues of subsite −1 of *Bl*Arap42B overlaid with galactose from *Bl*Gal42A (sphere). The Trp358 of *Bl*Arap42B, which is the smaller His in *Bl*Gal42A, is clashing with the C6-OH of galactose in complex with *Bl*Gal42A, explaining lack of activity towards β-galactosides of *Bl*Arap42B. **(C)** Active site residues of *Bl*Gal42A with paeonolide docked from *Bl*Arap42B shown as spheres. Both the side chains of Trp332 and Phe226 at subsite +1 and +2 sterically hinder the docking of paeonolide, in agreement with the lack of activity for *Bl*Gal42A.

FIGURE 4. *Bl*Arap42B defines a novel GH42 α-L-arabinopyranosidase subfamily.

The phylogenetic tree of GH42 enzymes, divides the family into the two distinct α-L-arabinopyranosidases (subfamily A) and β-galactosidases (subfamily G) subfamilies, respectively. Bifidobacterial GH42 enzymes (black) from the CAZy database reveals four distinct groups, with functional assignment for groups G2, G3 and G4. Group G1 was previously uncharacterized and multiple sequence alignment shows this group is diverse and comprises a unique Trp-X_{12}-Trp sequence motif as compared to the Phe-X_{12}-His sequence motif of group G2, G3 and G4. Approximately 200 non-identical sequences (grey) found throughout the bacterial kingdom from each of the two Trp-X_{12}-Trp and Phe-X_{12}-His sequence motifs were merged with the bifidobacterial GH42 sequences and previously characterized GH42 enzymes (grey) to yield the phylogenetic tree.

TABLE 1. Specific activity (U mg^{-1}) of *Bl*Arap42B from *B. animalis* subsp. *lactis* BL-04 (wild-type and mutants) and *Ri*Arap42B from *Roseburia intestinalis* M50/1 towards 5 mM *p*NP-substrates

Substrate	*Bl*Arap42B				*Ri*Arap42B
	Wild-type	W345F	W358H	W345F/ W358H	
*p*NP-α-L-Ara*p*	79.0 ± 0.7	2.7 ± 0.1	5.9 ± 0.1	0.1 ± 0.0	55.2 ± 2.0
*p*NP-β-D-Gal*p*	0.1 ± 0.0	n.d.	n.d.	n.d.	0.1 ± 0.0
*p*NP-β-D-Fuc*p*	1.9 ± 0.0	n.d.	0.7 ± 0.0	n.d.	2.6 ± 0.1

U mg^{-1} = μmol min^{-1} mg^{-1}. n.d. not detected (<0.05 U mg^{-1}). The activity was measured at 37 °C in 40 mM sodium citrate (pH 6.5) and 0.005% Triton X-100. All enzymes including the mutants showed no detectable activity towards *p*NP-β-D-Glc*p*, *p*NP-α-L-Ara*f*, and *p*NP-β-D-Xyl*p*.

TABLE 2. Kinetic parameters of *Bl*Arap42B on plant glycosides and *p*NP-α-L-Ara*p*

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
Paeonolide[a]	0.074 ± 0.007	240 ± 10	3,200 ± 200
Ginsenoside Rb2[a]	0.051 ± 0.018	79 ± 8	1,500 ± 400
*p*NP-α-L-Ara*p*[b]	3.4 ± 0.4	160 ± 10	46 ± 0.3

[a] Measured at 30 °C in 40 mM sodium citrate (pH 6.5).
[b] Measured at 37 °C in 40 mM sodium citrate (pH 6.5) and 0.005% Triton X-100.

TABLE 3. Data collection and refinement statistics for *Bl*Arap42B

	*Bl*Ara*p*42B
Data collection	
Beamline	KEK-PF BL5A
Resolution (Å)	50.00–2.00 (2.03–2.00)
Space group	$P4_12_12$
Unit cell parameters (Å)	$a = b = 177.95$, $c = 375.70$
No. of reflections	5,987,058
No. of unique reflections	400,874 (19,842)
Mean I/σ (I)	27.7 (3.8)
***R*-sym (%)**	12.7 (50.2)
Completeness (%)	100.0 (100.0)
Redundancy	14.9 (14.6)
Refinement	
Resolution (Å)	50.00–2.00
***R*-factor (%)**	15.5
***R*-free (%)**	19.1
No. of atoms:	
Protein atoms (non-hydrogen)	32,478
Water molecules	3,330
RMSD bond lengths (Å)	0.022
RMSD bond angles (°)	1.972
Average B-factors (Å^2):	
Protein atoms	26.3
Water	47.9
Ramachandran plot (%):	
Favored regions	96.0
Allowed regions	3.8
Disallowed regions	0.2
PDB code	5XB7

Numbers in parentheses are for the highest resolution shell.

Figure 1

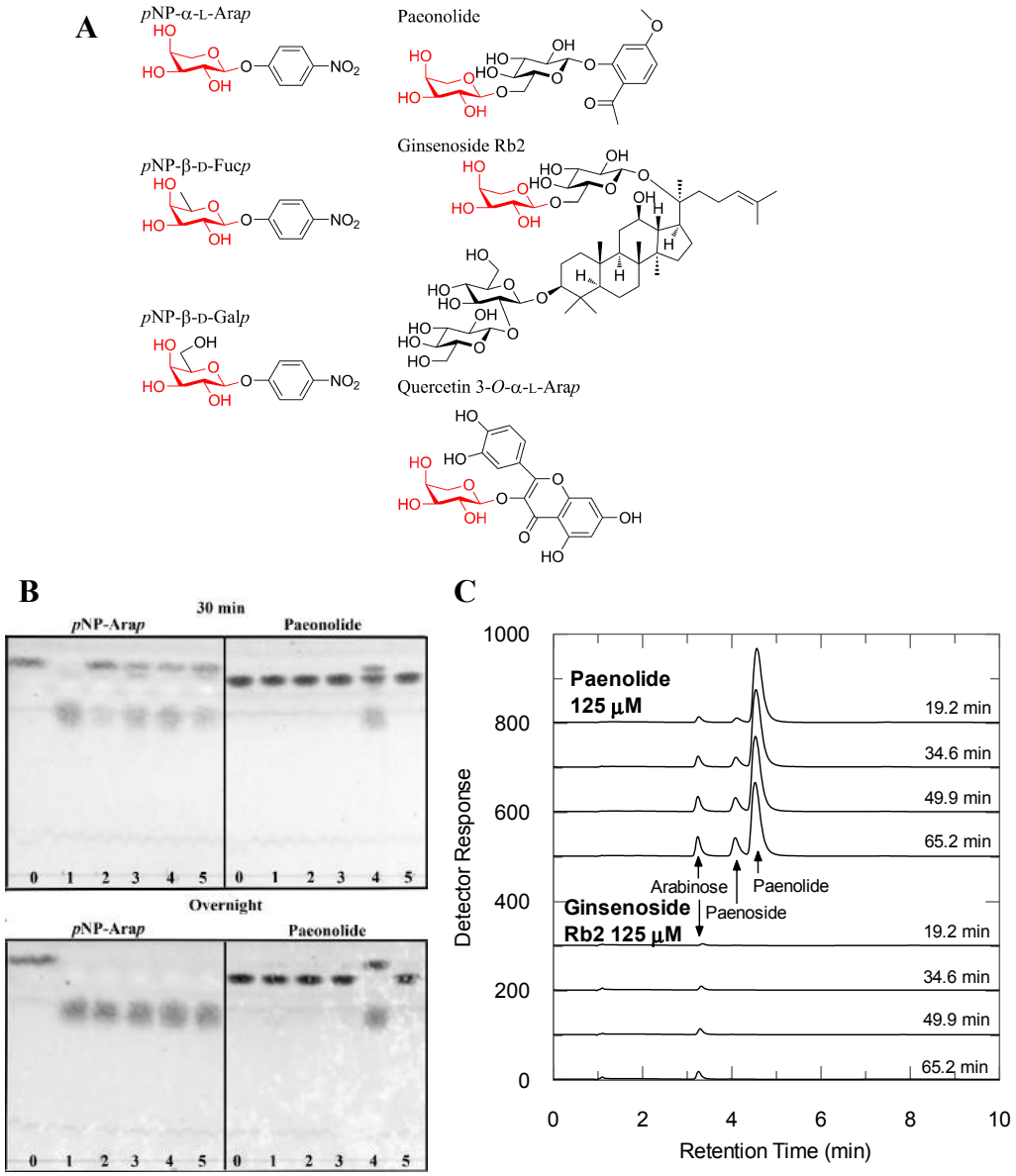


Figure 2

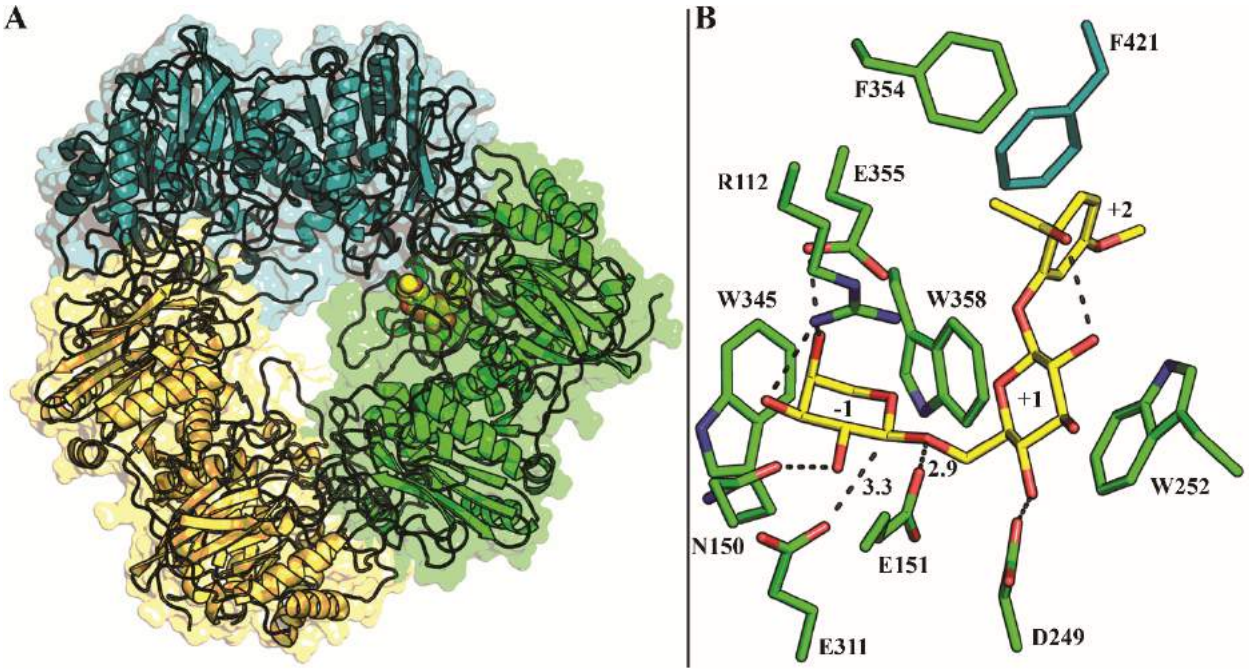


Figure 3

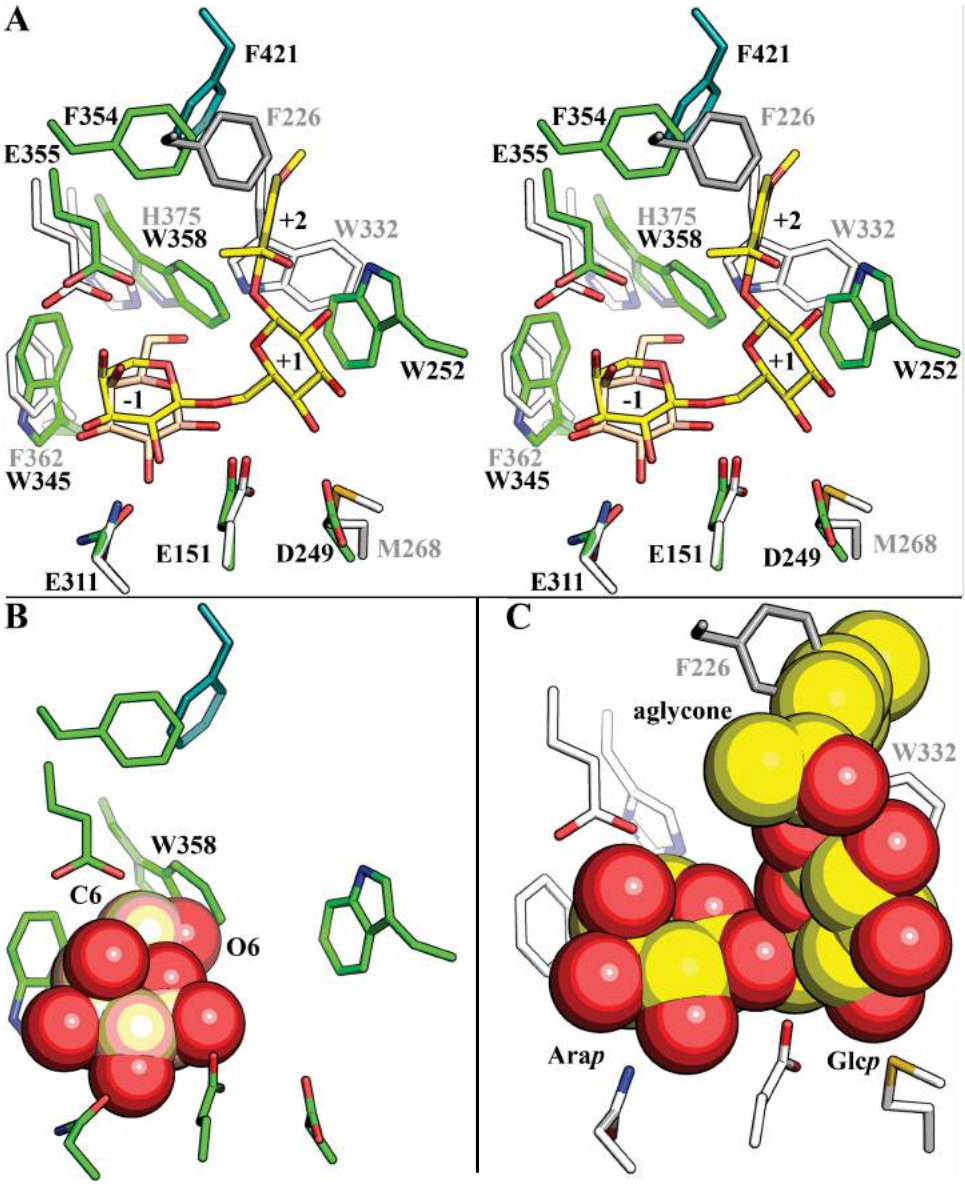
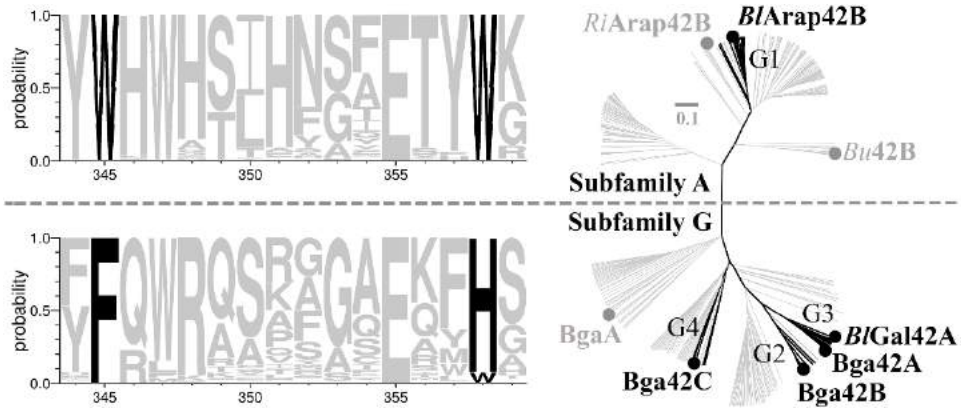


Figure 4



Discovery of α-L-arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42

Alexander Holm Viborg, Takane Katayama, Takatoshi Arakawa, Maher Abou Hachem, Leila Lo Leggio, Motomitsu Kitaoka, Birte Svensson and Shinya Fushinobu

J. Biol. Chem. published online October 23, 2017
